

# A TRIDENT SCHOLAR PROJECT REPORT

NO. 437

---

**Probe-Independent EEG Assessment of Mental Workload in Pilots**

by

Midshipman 1/C Michael K. Johnson, USN

---



UNITED STATES NAVAL ACADEMY  
ANNAPOLIS, MARYLAND

This document has been approved for public  
release and sale; its distribution is limited.

USNA-1531-2

U.S.N.A. --- Trident Scholar project report; no. 437 (2015)

**PROBE-INDEPENDENT EEG ASSESSMENT  
OF MENTAL WORKLOAD IN PILOTS**

by

Midshipman 1/C Michael K. Johnson  
United States Naval Academy  
Annapolis, Maryland

---

(signature)

Certification of Adviser Approval

Assistant Professor Justin A. Blanco  
Electrical and Computer Engineering Department

---

(signature)

---

(date)

Acceptance for the Trident Scholar Committee

Professor Maria J. Schroeder  
Associate Director of Midshipman Research

---

(signature)

---

(date)

USNA-1531-2

<b>REPORT DOCUMENTATION PAGE</b>				<i>Form Approved</i> <b>OMB No. 0704-0188</b>	
Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing this collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. <b>PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.</b>					
<b>1. REPORT DATE</b> (DD-MM-YYYY) 05-18-2015		<b>2. REPORT TYPE</b>		<b>3. DATES COVERED</b> (From - To)	
<b>4. TITLE AND SUBTITLE</b> Probe-Independent EEG Assessment of Mental Workload in Pilots				<b>5a. CONTRACT NUMBER</b>	
				<b>5b. GRANT NUMBER</b>	
				<b>5c. PROGRAM ELEMENT NUMBER</b>	
<b>6. AUTHOR(S)</b> Johnson, Michael Kamis				<b>5d. PROJECT NUMBER</b>	
				<b>5e. TASK NUMBER</b>	
				<b>5f. WORK UNIT NUMBER</b>	
<b>7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)</b>				<b>8. PERFORMING ORGANIZATION REPORT NUMBER</b>	
<b>9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)</b> U.S. Naval Academy Annapolis, MD 21402				<b>10. SPONSOR/MONITOR'S ACRONYM(S)</b>	
				<b>11. SPONSOR/MONITOR'S REPORT NUMBER(S)</b> Trident Scholar Report no. 437 (2015)	
<b>12. DISTRIBUTION / AVAILABILITY STATEMENT</b>  This document has been approved for public release; its distribution is UNLIMITED.					
<b>13. SUPPLEMENTARY NOTES</b>					
<b>14. ABSTRACT</b>  We developed probe-independent algorithms for classifying three levels of task-complexity based on 4-channel electroencephalographic (EEG) recordings during simulated flight. Using a library of 168 input features drawn from different signal processing application domains, we evaluated 10 different classifiers, using 10-fold cross-validation to estimate generalization performance. The best subsets of features for each subject yielded a median classification accuracy of 92.81%, with 100% accuracy in two subjects and greater than 70% in all 19 subjects. Generally, the EEG line length and linear discriminant analysis were among the most effective features and classifiers, respectively. However, to maximize performance, the feature set-classifier combinations should be chosen based on the individual. No single channel proved more valuable than another in predicting flight task-complexity, but fusing the information across channels improved performance in 18 of 19 subjects. Given the success we had in producing high classification accuracies without an auditory stimulus, we believe this algorithm may be useful in developing optimal equipment or training techniques to minimize mental workload, and/or to monitor the mental state of a pilot over the course of a mission.					
<b>15. SUBJECT TERMS</b> mental workload, electroencephalography, pilots, machine learning					
<b>16. SECURITY CLASSIFICATION OF:</b>			<b>17. LIMITATION OF ABSTRACT</b>	<b>18. NUMBER OF PAGES</b>  29	<b>19a. NAME OF RESPONSIBLE PERSON</b>
<b>a. REPORT</b>	<b>b. ABSTRACT</b>	<b>c. THIS PAGE</b>			<b>19b. TELEPHONE NUMBER</b> (include area code)

## Table of Contents

Abstract.....	1
Acknowledgements.....	2
Introduction.....	3
Methods.....	4
Results.....	7
Behavioral Task Performance Assessment.....	15
Conclusion.....	17
Appendix.....	18
References.....	26

## Abstract

We developed probe-independent algorithms for classifying three levels of task-complexity based on 4-channel electroencephalographic (EEG) recordings during simulated flight. Using a library of 168 input features drawn from different signal processing application domains, we evaluated 10 different classifiers, using 10-fold cross-validation to estimate generalization performance. The best subsets of features for each subject yielded a median classification accuracy of 92.81%, with 100% accuracy in two subjects and greater than 70% in all 19 subjects. Generally, the EEG line length and linear discriminant analysis were among the most effective features and classifiers, respectively. However, to maximize performance, the feature set-classifier combinations should be chosen based on the individual. No single channel proved more valuable than another in predicting flight task-complexity, but fusing the information across channels improved performance in 18 of 19 subjects. Given the success we had in producing high classification accuracies without an auditory stimulus, we believe this algorithm may be useful in developing optimal equipment or training techniques to minimize mental workload, and/or to monitor the mental state of a pilot over the course of a mission.

**Keywords:** Mental workload, electroencephalography, pilots, machine learning

### Acknowledgements

Special thanks to Dr. Brad Hatfield, Dr. Rodolphe Gentili, Hyuk Oh, and Kyle Jaquess from the University of Maryland in College Park, MD and Dr. Jessica Mohler, CAPT Mike Prevost, and CDR Hartley Postlethwaite from the United States Naval Academy in Annapolis, MD for their data collection efforts, and to Lockheed Martin Corporation for its partial funding support of those efforts.

## Introduction

Mental workload can be described as a ratio between task-complexity and a person's cognitive capacity to meet task demands [1]. This description captures the intuitive idea that mental workload depends both on external factors such as the objective difficulty of required tasks, and internal factors such as a person's past experiences and skill set. There is a growing body of research focused on developing quantitative methods to assess mental workload in order to improve the mental resiliency of people in high stress environments. Various metrics derived from physiological signals such as heart rate, blood pressure, galvanic skin response, and eye-gaze have been investigated as biomarkers of mental workload [2-4]. These signals have been used to distinguish mental workload levels with accuracies significantly better than chance, but there are still no widely accepted standards or commercial products for mental workload monitoring.

With recent improvements in the ease-of-use, reliability, and costs of portable electroencephalography (EEG) systems, there has been increasing interest in using brain signals to measure mental workload [5]. It is hypothesized that EEG offers a more direct assay of mental workload than other physiological biomarkers because of the proximity of EEG sensors to the neural substrates of cognitive stress [6]. A common method of using EEG to assess mental workload involves delivering an auditory probe to evoke event-related potentials (ERPs) such as the P300 wave [7-8] while a subject undergoes a cognitive challenge. It has been demonstrated that the amplitude of the P300 varies inversely with task-complexity [9]. Although measuring ERPs using auditory stimuli has been successful in evaluating mental workload in controlled settings, this approach may be less practical when the cognitive tasks themselves involve a strong auditory component, and hence the introduction of extraneous sounds could be disruptive. Therefore, an EEG-based metric of mental workload that exploits information from non-evoked "background" neural activity is desired.

The goal of this research was to develop an EEG-based algorithm to classify different levels of task-complexity that does not rely upon ERPs. By choosing subjects with a similar level of task-experience, we partially control for differences in the capacity to perform the experimental task and therefore use task-complexity as a surrogate for mental workload. As we were particularly interested in understanding the response of aircraft pilots to the cognitive demands imposed by their flight-missions, we used flight simulator tasks of varying challenge-level as our experimental paradigm. Furthermore, since pilots are typically in persistent radio or intercom communications via headset during flight, this also represents a scenario that would be particularly well-suited to a non-ERP-based index of cognitive workload. Signal processing methods were used to extract computational features from the EEG, and machine learning techniques were used to classify the data and assess algorithm performance.

## Method

EEG data were collected from 19 United States Naval Academy (Annapolis, MD, USA) midshipmen between the ages of 19 and 23, all with basic flight training, while they performed visuo-motor tasks in a flight simulator (Prepar3D® v1.4, Lockheed Martin Corporation, Orlando, FL, USA) under three levels of task-complexity. The three tasks were selected from predefined flight training exercises developed with advice from experienced United States Navy pilots and distinguished in challenge-level by differences in weather intensity and mission requirements. Specifically, the three tasks were: 1) Easy: maintain aircraft's current altitude (4000 ft), heading (180°), and airspeed (180 kn). The weather was defined by no clouds, precipitation, or wind, and unlimited visibility; 2) Medium: maintain the aircraft's current heading (180°), airspeed (180 kn), and a "wings-level" attitude, while continuously making altitude changes between 4000 and 3000 ft, with ascent and descent rates of 1000 feet per minute (fpm). The sky was completely overcast (1/16 mi of visibility), but there was no precipitation and no wind; and 3) Hard: maintain the aircraft's current airspeed (180 kn), while changing heading between 180 and 090° at a 15-degree angle of bank, ascending while turning right and descending while turning left at 1000 fpm. The sky was completely overcast as in the Medium task, with no precipitation, but with the presence of a moderate (16 kn) easterly wind. One trial per task difficulty was conducted, in random order, consisting of a 1-minute setup period followed by a 10-minute flight segment. Additionally, for use in a separate analysis, audible stimuli were administered to participants via ear-bud speakers with random inter-stimulus intervals between 6 and 30 seconds to evoke the P300 response. Since the goal of this work was to analyze background EEG only, data surrounding these stimuli were excluded using a procedure described below. Fig. 1 illustrates the experimental setup.



Fig. 1. An individual performs a flight task in the simulator while wearing an EEG cap. The three experimental tasks required subjects to operate a simulated T-6A Texan II SP2 United States Navy aircraft using the control stick, throttle, and rudder pedals.

Four active, gel-free electrodes were used to measure EEG signals from sites along the frontal (Fz), fronto-central (FCz), central (Cz), and parietal (Pz) midline, based on the International 10-20 System [10]. The EEG cap was connected to an amplifier with an online band-pass filter from 0.01 to 60 Hz (g.USBamp®, g.tec medical engineering GmbH, Schiedlberg, Austria) through a driver-interface box (g.SAHARAsys®, g.tec medical engineering GmbH, Schiedlberg, Austria). Electrode impedances were maintained below 5 kOhm during data acquisition. The right mastoid was used as ground for the system and the left ear as the hardware reference. Data were also collected from the right ear for later re-referencing. EEG were sampled at a rate of 512 Hz, re-referenced in an EEG analysis software (BrainVision Analyzer 2, Brain Products GmbH, Munich, Germany) to an average-ear montage, and digitally lowpass filtered (in forward and reverse to give zero phase response) with a Butterworth filter with a cutoff frequency of 50 Hz and 48 dB/octave rolloff.

Data were visually inspected for the presence of eye-blink and muscle-activity artifacts, and these segments (47% of all recorded data) were excluded from analysis. In addition, all data within a 600 ms window following the onset of an auditory stimulus were excluded in order to eliminate P300 responses. Fig. 2 shows a typical P300 waveform, generated by averaging 30 post-stimulus intervals in a single trial.

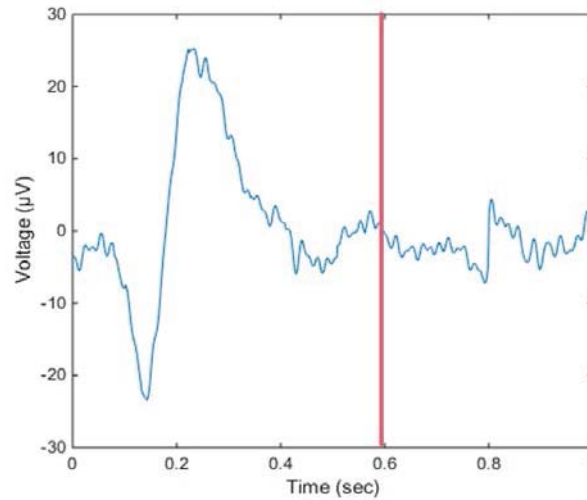


Fig. 2. A typical P300 waveform generated by averaging 30 post-stimulus intervals. The peak of the response is located at approximately 250 ms post-stimulus, and a return to baseline is seen by around 600 ms, supporting the choice of the 600 ms exclusion window.

The remaining data set was segmented into 1-second (512-sample) epochs for analysis, based primarily on a desire to have no more than 1-second lag in an envisioned online monitoring system. Any linear trends were removed from each segment by subtracting the least squares line of best fit. Feature vectors were then formed for each 1-second epoch by concatenating various subsets of 672 EEG measurements (168 measurements on each of 4 channels) based on signal processing techniques drawn from various application domains (e.g., biomedicine, speech

processing, and finance). The features collectively represent information extracted from the time, frequency, and wavelet domains, as well as parameters derived from information theory, nonlinear dynamics, and fractal geometry. A detailed list of features and their associated references are contained in the Appendix.

A total of 10 different classifiers were then trained to predict task-complexity based on EEG features. Broadly, the classification techniques used were: 1) k-Nearest Neighbors (kNN); 2) Linear Discriminant Analysis (LDA); 3) Quadratic Discriminant Analysis (QDA); 4) Naïve Bayes; 5) Decision Trees (with and without pruning); and 5) Support Vector Machines (SVM). Table I in the Appendix lists all classifiers and the values of their key properties and parameters. Test set error was estimated using 10-fold cross-validation, with percent correct classification used as the performance metric.

## Results

### A. Benchmark Analysis

Estimates of the EEG power in discrete frequency ranges that have known clinical significance commonly appear as input features in the brain-machine interface literature, including prior work on mental workload measurement [11-16]. As a preliminary benchmark for later comparison with more advanced processing, we evaluated the performance of a classifier that used only EEG spectral band power measurements as input features. We estimated the power spectral density (PSD) of each 1-second EEG segment using Welch's method (8 sections with 50% overlap; Hamming window applied to each section.) [17]. The average powers in each of the delta (1-4 Hz), theta (4-8 Hz), alpha (8-13 Hz), beta (13-30 Hz), and gamma (30-40 Hz) bands were then computed by integrating this PSD estimate over the corresponding frequency range. These five inputs, computed on each of the 4 recorded channels, formed the basis for the input feature vectors used in the classification tests described in this section.

Specifically, three sets of features were used in the benchmark analysis: (1) the average power in all five frequency bands; (2) the average power in the alpha, beta, and gamma frequency bands only [18-21]; and (3) the first, second, and third principal components of the average power in all five frequency bands. We performed principal components analysis on the unnormalized five-element feature data, reducing the number of dimensions to three, accounting for approximately 90% of the data variance on average, in an attempt to improve classifier performance by deemphasizing potentially irrelevant features.

Fig. 3 illustrates an example of one particular classifier selected for its ease of visualization: a linear discriminant analysis classifier using principal components as features for channel FCz in Subject 8. The decision boundaries are shown as black planes, and good class-separation is

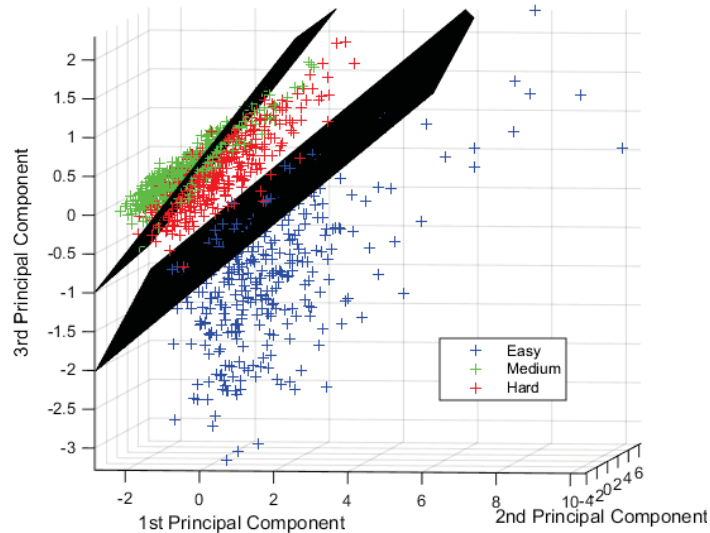


Fig. 3. Example of a linear discriminant analysis classifier (Subject 8, channel FCz). The input features were the first, second, and third principal components of the average power in the five frequency bands considered. The accuracy for this classifier was 82.91%.

evident, resulting in 82.91% classification accuracy. This subject and channel yielded some of the strongest results we observed in the benchmark analysis, suggesting that average powers in the selected frequency bands are plausible features for assessing mental workload in this subject.

Classification performance across subjects and channels is summarized in fig. 4. Nearly 70% of subjects (13 of 19) had at least one channel with a best-classification accuracy (i.e., highest of the ten classifiers considered; hereafter referred to simply as classification accuracy) above 50% using alpha, beta, and gamma power as inputs, while 9 subjects performed above 50% accuracy for all four channels. The lowest classification accuracy was 34.00%, while the highest was 87.35%. Chance performance is 33.33% using a null model of equiprobable classes, and 42.23% using a null model that assigns each observation to the most commonly occurring class in the training data.

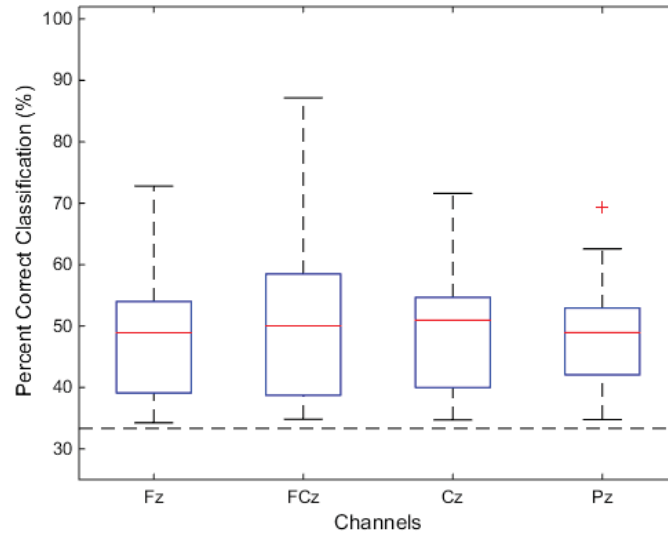


Fig.. 4. Each box represents the distribution of classification accuracies across all subjects for the best performing classifier-feature combination given the 3 feature sets tested in the benchmark analysis. The dotted line represents chance performance using a null model of equiprobable classes.

### B. All Features Analysis

Accuracy improved when all 168 features were used. Fig. 5 shows classification accuracy across all subjects for each channel. Accuracies above 50% were obtained for all subjects on at least one channel, and 13 of the 19 subjects had at least one channel above 75%. The lowest classification accuracy was now 47.32%, while the highest was 100%. Despite the improvement over the benchmark, results continued to exhibit considerable variability across subjects; the average interquartile range over channels was 21.00%.

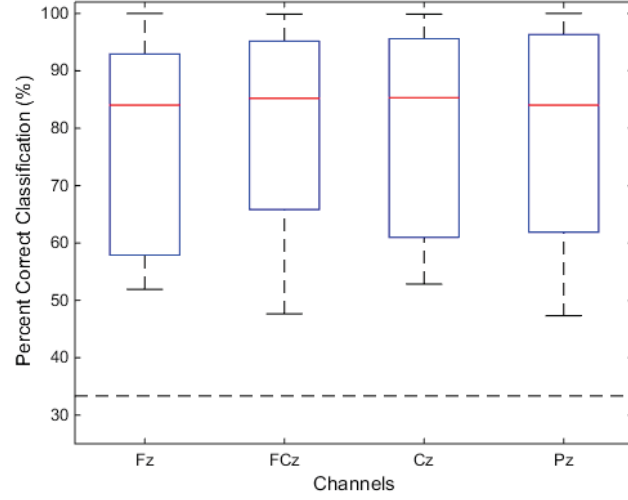


Fig. 5. Each box represents the distribution of classification accuracies across all subjects for the best performing classifier-feature combination using all 168 features. The dotted line represents chance performance using a null model of equiprobable classes.

### C. Forward Feature Selection

To mitigate overfitting, particularly in three subjects who had fewer than 300 observations, and to reduce the negative impacts of any features that might be redundant or unrelated to mental workload, we performed feature subset selection. Minimizing the number of features has the added benefit of reducing prediction delay and power consumption for trained classifiers, which is critical for the envisioned online realizations of the algorithm in portable hardware.

Since exhaustive search of the  $2^{168}$  possible feature-subsets to determine the optimal classification model for each channel is intractable, we used forward stepwise feature selection, which considers a much smaller search space of candidate models (at most  $1+168(168+1)/2 = 14197$  in our case). Forward stepwise feature selection begins with a model containing no features and, on each iteration, adds to the model the single feature that gives the greatest additional improvement in performance, in our case measured as 10-fold cross validated accuracy. The final subset was selected according to the one standard error rule [22-23]: the smallest subset with accuracy within one standard error of the mean of the best-performing subset. Fig. 6 shows how the number of features included in the model affects classification performance in one particular subject (subject 6) with relatively poor classification accuracy but a wide dynamic range of accuracies over the set of models considered. Since the final subset selected may depend on random partitions of the data (for 10-fold cross validation) made throughout the progression of the algorithm, forward feature selection was performed 250 times for each combination of subject and channel, yielding a total of  $19 \times 4 \times 250 = 19,000$  iterations.

Linear discriminant analysis was used for the feature selection process because LDA was chosen for 63.16% of classifiers during testing with all features and it requires less computational time per iteration than other classifiers. Table II lists the top ten features, ordered by the percentage of iterations on which those features were among the “optimal” subset selected. Measurements of

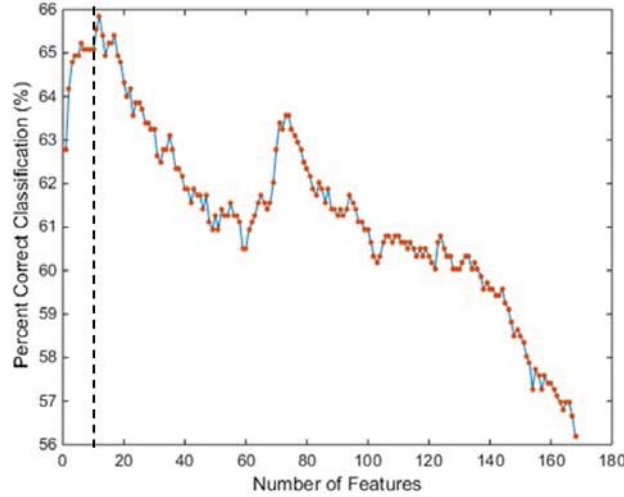


Fig. 6. Example of forward feature selection results. The vertical dotted line represents the number of features corresponding to the highest performance for this subject (subject 6) and channel (FCz).

line length and the number of peaks were the two most frequently selected at 88.23% and 64.61%, respectively and considerably higher than the third feature, the variance of the Teager energy operator, at 21.81%. Additionally, fig. 7 displays the distribution of the size of the best subset of features chosen. The average size of the feature subset across all iterations was 12.37 features.

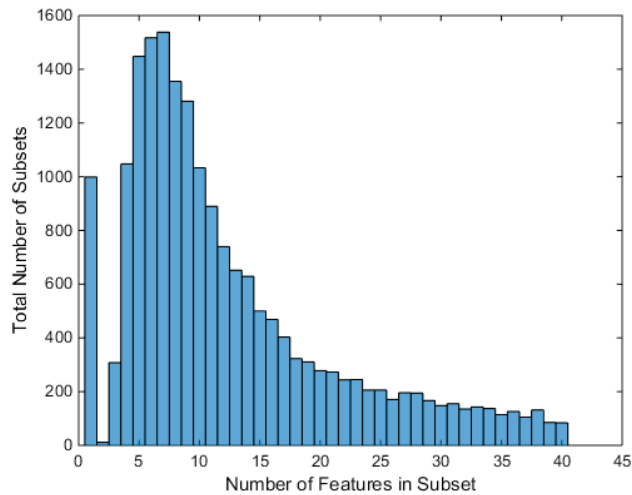


Fig. 7. Bar graph representation of the distribution of subset sizes across all iterations for the best features subset-classifier combinations found during the feature selection process.

TABLE II. FEATURE SELECTION

	Feature Names	Average Rank in Subset	Percent Selected (%)
1	Line Length - Time Series	2.12	89.12
2	Number of Peaks - Time Series	3.97	65.93
3	Variance - Teager Energy Operator	9.74	25.99
4	Variance of Power Spectral Density	10.60	22.52
5	Mean of Power Spectral Density - z-score	9.06	22.46
6	Mean - Teager Energy Operator - Frequency Modulated Component	8.73	20.57
7	Average Power - Beta	9.06	19.88
8	Kurtosis - Wavelet Decom. Coefficients - Daubechies4 - Gamma	4.38	18.72
9	Mean - Teager Energy Operator - Frequency Modulated Component - z-score	10.94	17.46
10	Hurst Exponent - Discrete Second Order Derivative	7.02	17.06

#### *D. Best Subset Analysis*

For each subject, the best performing subset of features was selected from the 1000 iterations (4 channels x 250 iterations = 1000) of forward feature selection, totaling nineteen subsets. Because we were interested in exploring whether a “universal” set of features with good generalization performance across subjects could be found, nineteen additional subsets were tested. These were created by considering the number of times each feature was selected for any of the 1000 iterations in a given subject; the most frequently occurring twenty features per subject formed the nineteen additional subsets. These 38 total subsets were then tested across all subjects with all classifiers. Fig. 8 displays the highest performing combination of feature subset of these 38 and classifier across all subjects and channels. All subjects performed with at least one channel above 65% accuracy, and 12 of the 19 subjects performed with at least one channel above 80%. The lowest classification accuracy was 56.57%, and the highest was maintained at 100%.

Additionally, of the total 2888 classifiers (19 subjects, 4 channels, and 38 subsets), the Linear Discriminant Analysis and k-Nearest Neighbor (cross validated for best nearest neighbor) classifiers were chosen as the highest performing classifier in approximately 60% and 20%,

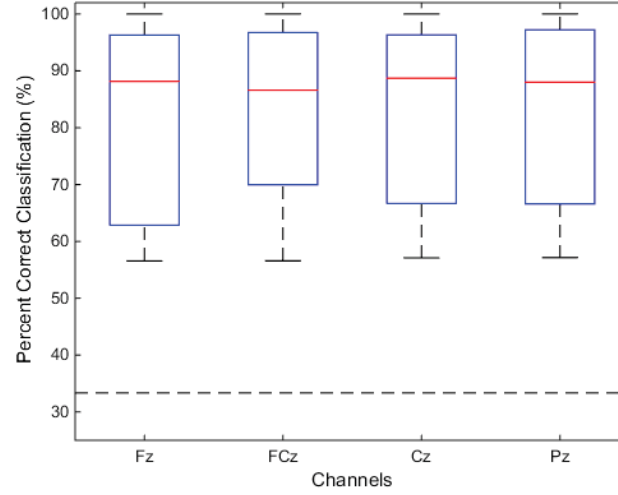


Fig. 8. Each box represents the distribution of classification accuracies across all subjects for the best performing classifier-feature combination using the 38 feature subsets described in the text. The dotted line represents chance performance using a null model of equiprobable classes.

respectively, of cases. Although this percentage suggests LDA may generally be the highest performing classifier, assurance of maximum performance demands testing all potential classifiers.

Fig. 9 compares the distributions of classification accuracies for channel FCz between models

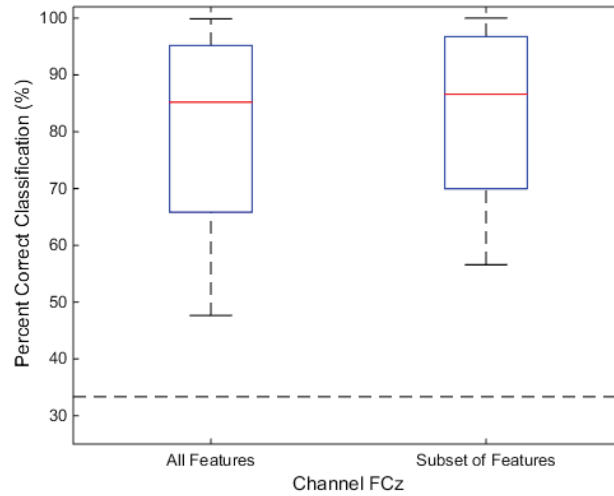


Fig. 9. The box on the left represents the distribution of classification accuracies across all subjects using all features as the input vector, while the box on the right represents the distribution of classification accuracies across all subjects using all 38 feature subsets for channel FCz. The dotted line represents chance performance using a null model of equiprobable classes.

that include all features and those that include the highest performing subset of features. The Wilcoxon Signed Rank test rejects the null hypothesis that the difference between paired samples has zero median ( $p < 0.001$ ), indicating a statistically significant improvement in classifier performance by using a subset of features rather than all the features.

#### *E. Combination of Channels Analysis*

We hypothesized that combining the information from all four channels in the input feature vector would improve classifier performance. To test this, all 168 features from each of the four channels were concatenated to form a single 672-element feature vector. Using forward feature selection with linear discriminant analysis, we identified the highest performing subset of features for each subject. These nineteen subsets were again tested across all subjects and classifiers to determine the highest performing combination of feature set and classifier. As illustrated in fig. 10, classification accuracy improved for 18 of the 19 subjects (one was already at 100%). The inclusion of all channels in the feature selection process allowed all subjects to perform above 70% accuracy, and 11 of the 19 subjects performed above 90% with two at 100%. Comparing fig. 10 with figs. 4, 5, and 8, it is evident that the algorithm achieved the highest performance when all channels were included in the classifier model, and a subset of features was subsequently chosen. Each subject maximized performance with a different combination of features and classifiers. Table III summarizes the results from the four stages of algorithm development across all subjects.

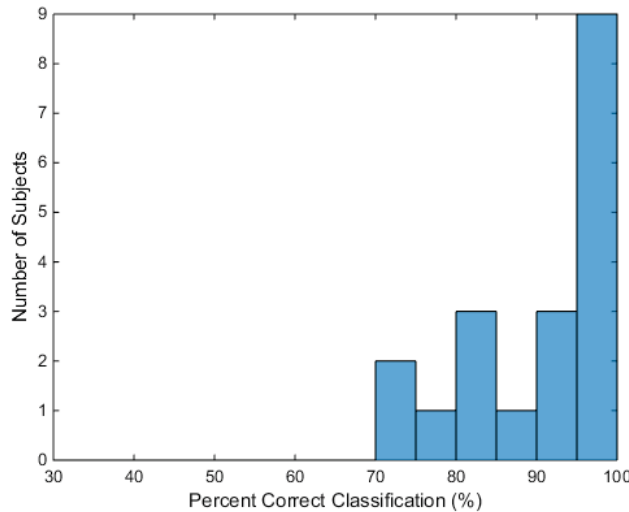


Fig. 10. Histogram estimate of the distribution of classification accuracies for the best feature subset-classifier combination when all channels are included in the model.

TABLE III: PROGRESSION OF ALGORITHM  
DEVELOPMENT PERCENT CORRECT  
CLASSIFICATION (%)

Subject	Phase 1	Phase 2	Phase 3	Phase 4
4	41.76	97.40	99.21	99.58
6	55.97	59.95	65.39	75.65
8	69.61	90.29	92.17	98.42
10	48.09	55.04	61.18	73.82
11	54.23	54.62	63.19	71.88
13	60.60	59.39	64.14	90.48
17	39.13	99.95	100.00	100.00
18	74.26	86.24	91.72	98.22
19	54.02	55.58	62.72	80.36
22	61.35	86.25	88.05	92.81
23	51.72	75.39	77.77	86.40
26	52.06	71.46	73.02	81.27
29	43.92	90.26	93.41	95.34
30	44.86	96.01	97.46	98.66
33	56.97	76.49	82.21	92.63
34	56.01	56.71	61.78	82.21
35	40.06	96.13	97.71	99.30
36	39.03	93.58	93.56	95.23
39	50.80	95.74	98.94	100.00

## Behavioral Task Performance Assessment

Regardless of the accuracy of predicting task complexity based on EEG characteristics, there persists the question of how tight the coupling is between task complexity and mental workload. To address this question, behavioral performance data for each subject, including measurements of air speed, heading, altitude, vertical air speed, rate of turn, and angle of bank, was analyzed. We hypothesized that subjects who showed poor behavioral performance on the easy flight task were cognitively overloaded from the outset of the experiment and therefore would not be expected to have EEG strongly modulated by task-complexity. Relatedly, we hypothesized that subjects who showed no decrease in originally strong behavioral performance as task complexity increased were not sufficiently challenged cognitively, and therefore would also not be expected to show strong EEG changes as task complexity varied.

Two methods were used to assess performance at each flight task. The methods were developed in discussions with professional pilots and flight instructors with expertise in assessing the performance of young pilots. The first was based on deviations from flight objectives outside the thresholds specified in the Methods section (e.g., maintain level flight at altitude of 4000 ft). Each time the subject exceeded these boundaries, it was counted as a “failure” and the total number of such failures per flight task was determined for each subject. For 13 of the 19 subjects, the number of failures increased as flight task-complexity increased. Fig. 11 compares the distributions of classification accuracies for subjects with an increase in failures as flight tasks become more difficult and those with no increase, using the all features-best classifier combination.

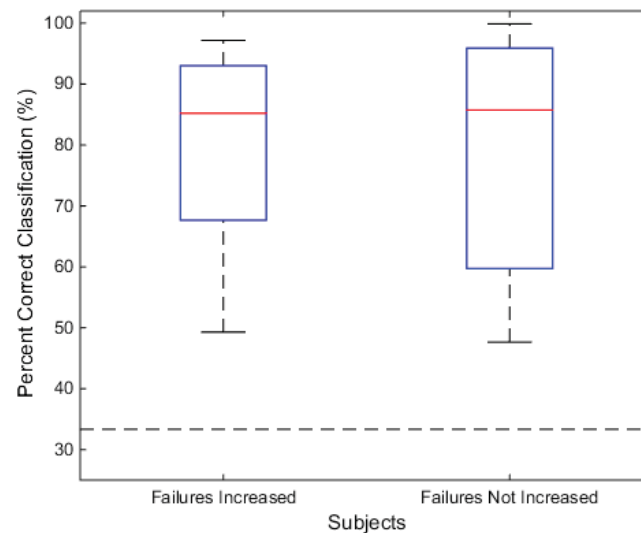


Fig. 11. Box plots showing the distribution of classification accuracies for subjects with and without behavioral performance decrements as flight task complexity increased, using the failure indicator metric of behavioral performance. There were no significant differences between the groups (Wilcoxon rank sum test,  $p = 0.765$ ). The dotted line represents chance performance using a null model of equiprobable classes.

The second method of measuring behavioral performance was an index developed by the University of Maryland that was scaled between 0, representing high performance, and 1, representing low performance [24]. Thirteen of the 19 subjects showed an increase in the behavioral performance index as task complexity increased (though not the identical set of 13 detected using the failure index discussed above). Fig. 12 shows classification results when the University of Maryland behavioral performance index is used to partition subjects.

The second performance assessment using University of Maryland's behavioral performance index does reflect a strong trend that subjects exhibiting an increase in performance index have higher classification rates. This supports the assumption that task complexity may be an appropriate surrogate for mental workload if "saturation effects" are appropriately accounted for.

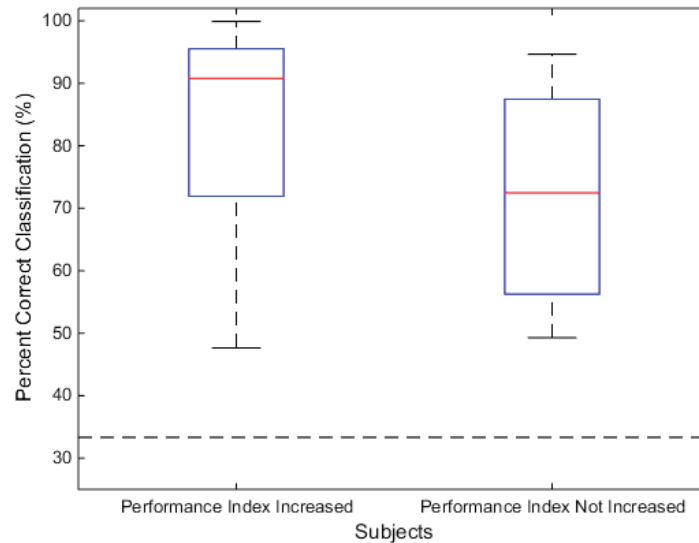


Fig. 12. Box plots showing the distribution of classification accuracies for subjects with and without behavioral performance decrements as flight task complexity increased, using the University of Maryland index of behavioral performance. The Wilcoxon Rank Sum test fails to reject the null hypothesis the two distributions are identical ( $p = 0.17912$ ), though there is a strong trend toward higher classification accuracies in the group showing behavioral performance decrements. The dotted line represents chance performance using a null model of equiprobable classes.

## Conclusion

The purpose of this research was to train a classifier to predict different levels of task complexity, and by extension, mental workload, using features extracted from background EEG collected during a flight simulator experiment. Classification accuracies above 70% were achieved for all subjects, with two subjects attaining 100%, but the best-performing combination of feature subset and classifier varied across subjects. Of note, classifiers were configured using typical but untuned parameters. Optimizing the latter could yield higher classification accuracies than already achieved. Generally, the best performing two features were line length and the number of peaks in a time segment, and the best performing classifier was linear discriminant analysis.

No single channel proved consistently better than another during classification. However, combining all channels into a single feature vector, followed by subset selection, improved performance for 18 of the 19 subjects (all subjects for whom improvement was possible). Practical online implementation of a classifier for mental workload predictions may be feasible given that we used EEG measurements in 1-second historical windows, but the impacts of classifier training and testing time, as well as feature computation time, need to be further studied. Additionally, the algorithm (e.g., LDA with a best-feature subset determined as described in this report) should be implemented on a microcontroller, DSP, or ASIC in a more portable system to test the feasibility of real-world implementation. We also hypothesize that incorporating non-EEG features, such as eye-tracking information, skin conductance, and heart rate variability, may improve classification performance.

We anticipate that the results of this study will guide future implementation of mental workload monitoring in operational environments such as those encountered by a military pilot in a cockpit. Given the success we had in producing high classification accuracies without an auditory stimulus, we have reason to believe this algorithm may be useful in developing optimal equipment or training techniques to minimize mental workload, and/or to monitor the mental state of a pilot over the course of a mission. The Navy will be able to use such an algorithm to enhance the mental resiliency of its warfighters, improving the overall safety, productivity, and performance of our nation's military.

## Appendix

TABLE I: LIST OF CLASSIFIERS TESTED AND KEY PROPERTIES AND PARAMETERS

Classifier	Value 1	Value 2	Value 3
1-Nearest Neighbor	Euclidean Distance	N/A	N/A
k-Nearest Neighbor	Euclidean Distance	k chosen based on cross validation of nearest neighbors from 3 to 51	N/A
Linear Discriminant Analysis	N/A	N/A	N/A
Quadratic Discriminant Analysis	N/A	N/A	N/A
Naïve Bayes	Distribution: Normal	N/A	N/A
Decision Trees (Unpruned)	Split Criterion: Gini's Diversity Index	N/A	N/A
Decision Trees (Pruned - CV)	Split Criterion: Gini's Diversity Index	Prune Criterion: Error	N/A
SVM (Radial Basis Function)	One vs. All	N/A	N/A
SVM (Linear)	One vs. All	N/A	N/A
SVM (Polynomial)	One vs. All	Order 2 or 3 chosen based on cross validation	N/A

### A. Mathematical Description of Select Features

The following are descriptions of the top three features chosen for subset selection indicated on Table II: line length, number of peaks, and the variance of the Teager energy operator (TEO). These are described in detail because of their frequency in determining the best subset of features. Line length (L) refers to a summation of the absolute value of the difference between two data points across the entire data set, as shown in Eqn. 1.

$$L = \sum_{n=0}^{N-1} |x[n] - x[n-1]| \quad (1)$$

The line length feature tends to be sensitive to signal amplitude and frequency variation, and was originally developed for use in detecting the onset of seizures [25]. The number of peaks counts the instances the signal changes from a positive trend to a negative trend. The Teager energy operator ( $\Psi$ ), originally developed to describe the behavior of energy in speech signals [26], is defined as:

$$\Psi[n] = x^2[n] - x[n-1]x[n+1] \quad (2)$$

The variance, a common statistic to describe the variability of data, of the Teager energy operator (TEO) is shown in Eqn. 3 and Eqn. 4.

$$\text{Var}(\Psi[n]) = \frac{1}{N} \sum_{n=0}^{N-1} (\Psi[n] - E[\Psi[n]])^2 \quad (3)$$

$$E[\Psi[n]] = \frac{1}{N} \sum_{n=0}^{N-1} \Psi[n] \quad (4)$$

### *B. Abbreviated Feature List*

The following list contains the shortened names of all 168 features.

1. Average Power – Delta Band
2. Average Power – Theta Band
3. Average Power – Alpha Band
4. Average Power – Beta Band
5. Average Power – Gamma Band
6. Average Power – Z-score – Delta Band
7. Average Power – Z-score – Theta Band
8. Average Power – Z-score – Alpha Band
9. Average Power – Z-score – Beta Band
10. Average Power – Z-score – Gamma Band
11. Mean of the Power Spectral Density (PSD)
12. Standard Deviation of PSD
13. Variance of PSD
14. Skewness of PSD
15. Kurtosis of PSD
16. Mean of the Power Spectral Density (PSD) - Z-score
17. Standard Deviation of PSD – Z-score
18. Variance of PSD – Z-score
19. Skewness of PSD – Z-score
20. Kurtosis of PSD – Z-score
21. Number of Zero Crossings – Time Series
22. Number of Zero Crossings – Time Series – Z-score
23. Number of Peaks – Time Series
24. Number of Peaks – Time Series – Z-score
25. Mean – Teager Energy Operator
26. Standard Deviation – Teager Energy Operator
27. Variance – Teager Energy Operator
28. Skewness – Teager Energy Operator
29. Kurtosis – Teager Energy Operator
30. Mean – Teager Energy Operator - Z-score
31. Standard Deviation – Teager Energy Operator - Z-score
32. Variance – Teager Energy Operator - Z-score
33. Skewness – Teager Energy Operator - Z-score
34. Kurtosis – Teager Energy Operator – Z-score
35. Mean – Teager Energy Operator – Frequency Modulated Component
36. Standard Deviation – Teager Energy Operator– Frequency Modulated Component
37. Variance – Teager Energy Operator– Frequency Modulated Component
38. Skewness – Teager Energy Operator– Frequency Modulated Component
39. Kurtosis – Teager Energy Operator– Frequency Modulated Component
40. Mean – Teager Energy Operator– Frequency Modulated Component - Z-score
41. Standard Deviation – Teager Energy Operator– Frequency Modulated Component - Z-score

42. Variance – Teager Energy Operator– Frequency Modulated Component - Z-score
43. Skewness – Teager Energy Operator– Frequency Modulated Component - Z-score
44. Kurtosis – Teager Energy Operator– Frequency Modulated Component – Z-score
45. Line Length – Time Series
46. Line Length – Time Series – Z-score
47. Hurst Exponent – Discrete Second Order Derivative
48. Hurst Exponent – Wavelet Based Adaptation
49. Hurst Exponent – Rescaled Range
50. Hurst Exponent – Discrete Second Order Derivative - Z-score
51. Hurst Exponent – Wavelet Based Adaptation - Z-score
52. Hurst Exponent – Rescaled Range – Z-score
53. Mean – Wavelet Decom. Coefficients – Db4 – Delta
54. Standard Deviation – Wavelet Decom. Coefficients – Db4 – Delta
55. Variance – Wavelet Decom. Coefficients – Db4 – Delta
56. Skewness – Wavelet Decom. Coefficients – Db4 – Delta
57. Kurtosis – Wavelet Decom. Coefficients – Db4 – Delta
58. Max – Wavelet Decom. Coefficients – Db4 – Delta
59. Min – Wavelet Decom. Coefficients – Db4 – Delta
60. Mean – Wavelet Decom. Coefficients – Db4 – Theta
61. Standard Deviation – Wavelet Decom. Coefficients – Db4 – Theta
62. Variance – Wavelet Decom. Coefficients – Db4 – Theta
63. Skewness – Wavelet Decom. Coefficients – Db4 – Theta
64. Kurtosis – Wavelet Decom. Coefficients – Db4 – Theta
65. Max – Wavelet Decom. Coefficients – Db4 – Theta
66. Min – Wavelet Decom. Coefficients – Db4 – Theta
67. Mean – Wavelet Decom. Coefficients – Db4 – Alpha
68. Standard Deviation – Wavelet Decom. Coefficients – Db4 – Alpha
69. Variance – Wavelet Decom. Coefficients – Db4 – Alpha
70. Skewness – Wavelet Decom. Coefficients – Db4 – Alpha
71. Kurtosis – Wavelet Decom. Coefficients – Db4 – Alpha
72. Max – Wavelet Decom. Coefficients – Db4 – Alpha
73. Min – Wavelet Decom. Coefficients – Db4 – Alpha
74. Mean – Wavelet Decom. Coefficients – Db4 – Beta
75. Standard Deviation – Wavelet Decom. Coefficients – Db4 – Beta
76. Variance – Wavelet Decom. Coefficients – Db4 – Beta
77. Skewness – Wavelet Decom. Coefficients – Db4 – Beta
78. Kurtosis – Wavelet Decom. Coefficients – Db4 – Beta
79. Max – Wavelet Decom. Coefficients – Db4 – Beta
80. Min – Wavelet Decom. Coefficients – Db4 – Beta
81. Mean – Wavelet Decom. Coefficients – Db4 – Gamma
82. Standard Deviation – Wavelet Decom. Coefficients – Db4 – Gamma
83. Variance – Wavelet Decom. Coefficients – Db4 – Gamma
84. Skewness – Wavelet Decom. Coefficients – Db4 – Gamma
85. Kurtosis – Wavelet Decom. Coefficients – Db4 – Gamma
86. Max – Wavelet Decom. Coefficients – Db4 – Gamma
87. Min – Wavelet Decom. Coefficients – Db4 – Gamma

88. Mean – Wavelet Decom. Coefficients – Db4 – Delta - Z-score
89. Standard Deviation – Wavelet Decom. Coefficients – Db4 – Delta - Z-score
90. Variance – Wavelet Decom. Coefficients – Db4 – Delta - Z-score
91. Skewness – Wavelet Decom. Coefficients – Db4 – Delta - Z-score
92. Kurtosis – Wavelet Decom. Coefficients – Db4 – Delta - Z-score
93. Max – Wavelet Decom. Coefficients – Db4 – Delta - Z-score
94. Min – Wavelet Decom. Coefficients – Db4 – Delta - Z-score
95. Mean – Wavelet Decom. Coefficients – Db4 – Theta - Z-score
96. Standard Deviation – Wavelet Decom. Coefficients – Db4 – Theta - Z-score
97. Variance – Wavelet Decom. Coefficients – Db4 – Theta - Z-score
98. Skewness – Wavelet Decom. Coefficients – Db4 – Theta - Z-score
99. Kurtosis – Wavelet Decom. Coefficients – Db4 – Theta - Z-score
100. Max – Wavelet Decom. Coefficients – Db4 – Theta - Z-score
101. Min – Wavelet Decom. Coefficients – Db4 – Theta - Z-score
102. Mean – Wavelet Decom. Coefficients – Db4 – Alpha - Z-score
103. Standard Deviation – Wavelet Decom. Coefficients – Db4 – Alpha - Z-score
104. Variance – Wavelet Decom. Coefficients – Db4 – Alpha - Z-score
105. Skewness – Wavelet Decom. Coefficients – Db4 – Alpha - Z-score
106. Kurtosis – Wavelet Decom. Coefficients – Db4 – Alpha - Z-score
107. Max – Wavelet Decom. Coefficients – Db4 – Alpha - Z-score
108. Min – Wavelet Decom. Coefficients – Db4 – Alpha - Z-score
109. Mean – Wavelet Decom. Coefficients – Db4 – Beta - Z-score
110. Standard Deviation – Wavelet Decom. Coefficients – Db4 – Beta - Z-score
111. Variance – Wavelet Decom. Coefficients – Db4 – Beta - Z-score
112. Skewness – Wavelet Decom. Coefficients – Db4 – Beta - Z-score
113. Kurtosis – Wavelet Decom. Coefficients – Db4 – Beta - Z-score
114. Max – Wavelet Decom. Coefficients – Db4 – Beta - Z-score
115. Min – Wavelet Decom. Coefficients – Db4 – Beta - Z-score
116. Mean – Wavelet Decom. Coefficients – Db4 – Gamma - Z-score
117. Standard Deviation – Wavelet Decom. Coefficients – Db4 – Gamma - Z-score
118. Variance – Wavelet Decom. Coefficients – Db4 – Gamma - Z-score
119. Skewness – Wavelet Decom. Coefficients – Db4 – Gamma - Z-score
120. Kurtosis – Wavelet Decom. Coefficients – Db4 – Gamma - Z-score
121. Max – Wavelet Decom. Coefficients – Db4 – Gamma - Z-score
122. Min – Wavelet Decom. Coefficients – Db4 – Gamma - Z-score
123. Energy – Wavelet Decom. Coefficients – Db4 – Delta
124. Energy – Wavelet Decom. Coefficients – Db4 - Theta
125. Energy – Wavelet Decom. Coefficients – Db4 - Alpha
126. Energy – Wavelet Decom. Coefficients – Db4 - Beta
127. Energy – Wavelet Decom. Coefficients – Db4 – Gamma
128. Energy – Wavelet Decom. Coefficients – Db4 – Delta - Z-score
129. Energy – Wavelet Decom. Coefficients – Db4 - Theta - Z-score
130. Energy – Wavelet Decom. Coefficients – Db4 - Alpha - Z-score
131. Energy – Wavelet Decom. Coefficients – Db4 - Beta - Z-score
132. Energy – Wavelet Decom. Coefficients – Db4 – Gamma - Z-score
133. Recoursing Energy Efficiency – Wavelet Decom. Coefficients – Db4 – Delta

134. Recoursing Energy Efficiency – Wavelet Decom. Coefficients – Db4 - Theta
135. Recoursing Energy Efficiency – Wavelet Decom. Coefficients – Db4 - Alpha
136. Recoursing Energy Efficiency – Wavelet Decom. Coefficients – Db4 - Beta
137. Recoursing Energy Efficiency – Wavelet Decom. Coefficients – Db4 – Gamma
138. Recoursing Energy Efficiency – Wavelet Decom. Coefficients – Db4 – Delta - Z-score
139. Recoursing Energy Efficiency – Wavelet Decom. Coefficients – Db4 - Theta - Z-score
140. Recoursing Energy Efficiency – Wavelet Decom. Coefficients – Db4 - Alpha - Z-score
141. Recoursing Energy Efficiency – Wavelet Decom. Coefficients – Db4 - Beta - Z-score
142. Recoursing Energy Efficiency – Wavelet Decom. Coefficients – Db4 - Gamma - Z-score
143. Logarithmic REE– Wavelet Decom. Coefficients – Db4 – Delta
144. Logarithmic REE– Wavelet Decom. Coefficients – Db4 – Theta
145. Logarithmic REE– Wavelet Decom. Coefficients – Db4 – Alpha
146. Logarithmic REE– Wavelet Decom. Coefficients – Db4 – Beta
147. Logarithmic REE– Wavelet Decom. Coefficients – Db4 – Gamma
148. Logarithmic REE– Wavelet Decom. Coefficients – Db4 – Delta - Z-score
149. Logarithmic REE– Wavelet Decom. Coefficients – Db4 – Theta - Z-score
150. Logarithmic REE– Wavelet Decom. Coefficients – Db4 – Alpha - Z-score
151. Logarithmic REE– Wavelet Decom. Coefficients – Db4 – Beta - Z-score
152. Logarithmic REE– Wavelet Decom. Coefficients – Db4 – Gamma - Z-score
153. Absolute LREE – Wavelet Decom. Coefficients – Db4 – Delta
154. Absolute LREE – Wavelet Decom. Coefficients – Db4 – Theta
155. Absolute LREE – Wavelet Decom. Coefficients – Db4 – Alpha
156. Absolute LREE – Wavelet Decom. Coefficients – Db4 – Beta
157. Absolute LREE – Wavelet Decom. Coefficients – Db4 – Gamma
158. Absolute LREE – Wavelet Decom. Coefficients – Db4 – Delta - Z-score
159. Absolute LREE – Wavelet Decom. Coefficients – Db4 – Theta - Z-score
160. Absolute LREE – Wavelet Decom. Coefficients – Db4 – Alpha - Z-score
161. Absolute LREE – Wavelet Decom. Coefficients – Db4 – Beta - Z-score
162. Absolute LREE – Wavelet Decom. Coefficients – Db4 – Gamma – Z-score
163. Correlation Dimension – Grassberger-Proccacia Algorithm
164. Correlation Dimension – Grassberger-Proccacia Algorithm – Z-score
165. Entropy – Time Series
166. Entropy – Time Series – Z-score
167. Optimum Minimum Time Delay – Phase Space Reconstruction – Time Series
168. Optimum Minimum Time Delay – Phase Space Reconstruction – Time Series - Z-score

### *C. Glossary of Terms*

The following is a glossary of terms found in the description of features:

*Absolute LREE:*

Absolute value of the Logarithmic Recoursing Energy Efficiency [11]

*Alpha Band:*

Frequency ranges between 8 – 13 Hz

*Beta Band:*

Frequency ranges between 13 – 30 Hz

*Correlation Dimension:*

Measure of the dimensionality of a set of data [33]

*Db4:*

Indicates the Daubechies 4 wavelet

*Delta Band:*

Frequency ranges between 1 – 4 Hz

*Entropy:*

Generally described as the amount of disorder in a system, entropy is generalized as the amount of information stored in a general probability distribution [34-35]

*Frequency Modulated Component of Teager Energy Operator (TEO):*

Nonlinear combinations of instantaneous signal outputs from TEO to separate its output energy product into its frequency component [29]; originally used in speech processing

*Gamma Band:*

Frequency ranges between 30 – 40 Hz

*Grassberger-Procaccia Algorithm:*

A method for estimating the correlation dimension [33]

*Hurst Exponent:*

Measure of the smoothness of the fractal dimension of the time series where fractal dimension refers to a measurement of the complexity of a signal [30]

*Kurtosis:*

Describes the shape of a probability distribution [27]

*Logarithmic Recoursing Energy Efficiency:*

Ten base logarithm of the REE [11]

Mean:

Average value of data set

Optimum Minimum Time Delay:

Optimal time delay to use for phase space reconstruction [36]

Phase Space Reconstruction:

Reconstruction of signal into phase space rather than in the time domain to view the dynamics of the system [36]

Power Spectral Density (PSD):

Describes how the power of a signal is distributed across a range of frequencies [28]

Recoursing Energy Efficiency (REE):

Measures the percentage of energy calculated from the wavelet decomposition coefficients of a signal within a sub-band [32]

Skewness:

Measure of the asymmetry of the probability distribution of a real-valued random variable about its mean [27]

Standard Deviation:

Measure used to determine the distribution of a set of data

Theta Band:

Frequency ranges between 4 – 8 Hz

Wavelet Decomposition Coefficients:

Refers to the coefficients calculated from the decomposition of signal into a series of basis functions called wavelets [31]; these coefficients represent dilations, contractions, and shifts

Variance:

Measure of the separation of data in a set; low variance means data points tend to be close, while high variance means the data is spread out

Z-score:

Refers to the standardization of the data to have a mean of 0 and a standard deviation of 1

## References

- [1] B. H. Kantowitz, "Human Factors Psychology 3. Mental Workload," *Advances in Psychology*, vol. 47, pp. 81 – 121, 1987.
- [2] T.C Hankins, G.F Wilson, "A comparison of heart rate, eye activity, EEG, and subjective measures of pilot mental workload during flight," *Aviation, Space, and Environmental Medicine.*, vol. 69, no. 4, pp. 360, 1998.
- [3] P. Besson, C. Bourdin, L. Bringoux, E. Dousset, C. Maïano, T. Marqueste, D. R. Mestre, S. Gaetan, J. Baudry, and J. Vercher, "Effectiveness of Physiological and Psychological Features to Estimate Helicopter Pilots' Workload: A Bayesian Network Approach," *IEEE Trans. Intell. Transp. Syst.*, vol. 14, no. 4, pp. 1872-1881, 2013.
- [4] W. Li, F. Chiu, K. Wu, and B. District, "The Evaluation of Pilots Performance and Mental Workload by Eye Movement," *Proceedings of the 30th European Association for Aviation Psychology Conference*, 2012, pp. 24–28.
- [5] T. K. Calibo, J. A. Blanco, and S. L. Firebaugh, "Cognitive stress recognition," *IEEE International Instrumentation and Measurement Technology Conference*, pp. 1471–1475, 2013.
- [6] A. Uetake, A. Murata, "Assessment of mental fatigue during VDT task using event-related potential," *Proceedings of the IEEE International Workshop on Robot and Human Interactive Communication*, 2000.
- [7] J. Song, D. Zhuang, G. Song, and X. Wanyan, "Pilot Mental Workload Measurement and Evaluation under Dual Task," pp. 809–812, 2011.
- [8] I. Käthner, S. C. Wriessnegger, G. R. Müller-Putz, A. Kübler, and S. Halder, "Effects of mental workload and fatigue on the P300, alpha and theta band power during operation of an ERP (P300) brain-computer interface.," *Biol. Psychol.*, vol. 102, pp. 118–29, Oct. 2014.
- [9] M. W. Miller, J. C. Rietschel, C. G. McDonald, and B. D. Hatfield, "A novel approach to the physiological measurement of mental workload.," *Int. J. Psychophysiol.*, vol. 80, no. 1, pp. 75–8, Apr. 2011.
- [10] V. Jurcak, D. Tsuzuki, and I. Dan, "10/20, 10/10, and 10/5 Systems Revisited: Their Validity As Relative Head-Surface-Based Positioning Systems," *NeuroImage*, vol. 34, no. 4, pp. 1600–11, Feb. 2007.
- [11] M. Murugappan, "Classification of human emotion from EEG using discrete wavelet transform," *J. Biomed. Sci. Eng.*, vol. 03, no. 04, pp. 390–396, 2010.
- [12] M. A. Hogervorst, A. Brouwer, and J. B. F. Van Erp, "Combining and comparing EEG, peripheral physiology and eye-related measures for the assessment of mental workload," *Front. Neurosci.*, vol. 8, pp. 1–14, Oct. 2014.

- [13] K. Brookhuis and D. de Waard, "Monitoring drivers' mental workload in driving simulators using physiological measures," *Accid. Anal. Prev.*, no. 42, pp. 898–903, 2010.
- [14] H. Petkar, R. Yadav, T. A. Nguyen, Y. Zeng, and S. Dande, "A Pilot Study to Assess Designer's Mental Stress using Eye Caze System and Electroencephalogram," *Proc. ASME Int. Des. Eng. Tech. Conf. Comput. Inf. Eng. Conf.*, pp. 1–11, 2009.
- [15] D. Dasari, G. Shou, and L. Ding, "Investigation of independent components based EEG metrics for mental fatigue in simulated ATC task," *Neural Eng.*, pp. 6–8, 2013.
- [16] E. B. J. Coffey, a.-M. Brouwer, and J. B. F. van Erp, "Measuring workload using a combination of electroencephalography and near infrared spectroscopy," *Proc. Hum. Factors Ergon. Soc. Annu. Meet.*, vol. 56, no. 1, pp. 1822–1826, Oct. 2012.
- [17] Welch, P.D., "A Fixed-point Fast Fourier Transform for the Estimation of Power Spectra," *IEEE Trans. Circuit Theory*, Vol.15, pp. 70-73, June 1970.
- [18] T. McMahan, T. D. Parsons, and I. Parberry, "Modality Specific Assessment of Video Game Player's Cognitive Workload Using Off-the-Shelf Electroencephalographic Technologies," 2014.
- [19] J. C. Rietschel, M. W. Miller, R. J. Gentili, R. N. Goodman, C. G. McDonald, and B. D. Hatfield, "Cerebral-cortical networking and activation increase as a function of cognitive-motor task difficulty," *Biol. Psychol.*, vol. 90, no. 2, pp. 127–33, May. 2012.
- [20] G. Borghini, L. Astolfi, G. Vecchiato, D. Mattia, and F. Babiloni, "Measuring neurophysiological signals in aircraft pilots and car drivers for the assessment of mental workload, fatigue and drowsiness.," *Neurosci. Biobehav. Rev.*, pp. 1–18, Oct. 2012.
- [21] M. M. Shaker, "EEG waves classifier using wavelet transform and Fourier transform," *World Academy of Science, Engineering and Technology*, vol. 3, pp. 723–728, 2007.
- [22] L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone, *Classification and Regression Trees*, 1984.
- [23] R. Tibshirani, G. Walther, and T. Hastie, "Estimating the number of clusters in a data set via the gap statistic," *J. R. Stat. Soc. Ser. B*, vol. 63, no. 2, pp. 411–423, 2001.
- [24] B. B. D. Hatfield, R. J. Gentili, J. C. Rietschel, L. Lo, H. Oh, K. J. Jaquess, M. Miller, and Y. Y. Tan, "Objective Assessment of Cognitive Workload & Attentional Reserve in Pilots during Varying Degrees of Task Difficulty," 2013.
- [25] P. Olsen, "Automatic detection of seizures using electroencephalographic signals," 1992.

- [26] J. F. Kaiser, "On a Simple Algorithm to Calculate the energy of a Signal," in *Acoustics, Speech, and Signal Processing*, 1988, pp. 381–384.
- [27] Y. Maruyama, "Measures of Multivariate Skewness and Kurtosis with Principal Components," *Japanese J. Appl. Stat.*, vol. 36, no. 3, pp. 139–145, 2007.
- [28] J. Zhou, "EEG Data Analysis, Feature Extraction and Classifiers," 2011.
- [29] P. Maragos, J. F. Kaiser, and T. F. Quatieri, "Energy Separation in Signal Modulations with Application to Speech Analysis," *IEEE Trans. Signal Process.*, vol. 41, no. 10, pp. 3024 – 3051, 1993.
- [30] M. Kale and F. Butar, "Fractal Analysis of Time Series and Distribution Properties of Hurst Exponent," *J. Math. Sci. Math. Educ.*, vol. 5, no. 1, pp. 8–19.
- [31] M. Akin, "Comparison of wavelet transform and FFT methods in the analysis of EEG signals," *J. Med. Syst.*, vol. 26, no. 3, pp. 241–7, Jun. 2002.
- [32] M. Murugappan, M. Rizon, R. Nagarajan, S. Yaacob, D. Hazry, and I. Zunaidi, "Time-Frequency Analysis of EEG Signals for Human Emotion Detection," *4th Kuala Lumpur Int. Conf. Biomed. Eng. 2008 IFMBE*, pp. 262–265, 2008.
- [33] J. Theiler, "Efficient algorithm for estimating the correlation dimension from a set of discrete points," *Physical Review A*, vol. 36, pp. 4456–4462, 1987.
- [34] H. Kantz and T. Schreiber, "Nonlinear Time Series Analysis," *Technometrics*, vol. 47, p. 369, 2004.
- [35] N. Kannathal, S. K. Puthusserypady, and L. Choo Min, "Complex dynamics of epileptic EEG," *Conf. Proc. IEEE Eng. Med. Biol. Soc.*, vol. 1, pp. 604–7, Jan. 2004.
- [36] M. B. Kennel, R. Brown, and H. D. I. Abarbanel, "Determining embedding dimension for phase-space reconstruction using a geometrical construction," *Physical Review A*, vol. 45, pp. 3403–3411, 1992.